Peter Klimczak; Isabel Kusche; Constanze Tschöpe; Matthias Wolff

Menschliche und maschinelle Entscheidungsrationalität. Zur Kontrolle und Akzeptanz Künstlicher Intelligenz

2019

https://doi.org/10.25969/mediarep/12631

Veröffentlichungsversion / published version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Klimczak, Peter; Kusche, Isabel; Tschöpe, Constanze; Wolff, Matthias: Menschliche und maschinelle Entscheidungsrationalität. Zur Kontrolle und Akzeptanz Künstlicher Intelligenz. In: *Zeitschrift für Medienwissenschaft*. Heft 21: Künstliche Intelligenzen, Jg. 11 (2019), Nr. 2, S. 39–45. DOI: https://doi.org/10.25969/mediarep/12631.

PETER KLIMCZAK / ISABEL KUSCHE /
CONSTANZE TSCHÖPE / MATTHIAS WOLFF

Menschliche und maschinelle Entscheidungsrationalität

Zur Kontrolle und Akzeptanz Künstlicher Intelligenz

I. Intelligente Maschinen

Der Begriff der «Künstlichen Intelligenz» (KI) wurde wesentlich auf der sogenannten Dartmouth Conference (1956) geprägt. John McCarthy, einer der Organisatoren, schrieb dazu:

> The study [of artificial intelligence] is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.[1]

Die Lösung dieser Aufgaben schien damals in naher Zukunft zu liegen. Die bisherige Entwicklung des Fachgebiets zeigt jedoch, dass diese Annahme zu optimistisch war.[2] Heute existieren leistungsfähige rechentechnische oder elektronische Nachbildungen von Intelligenzleistungen, z. B. dem Erkennen und Verstehen von Sprache und Objekten oder der Fähigkeit zum Lernen, zur Anpassung sowie zum zielführenden Handeln unter Unsicherheit. Die technische Nachbildung *aller* Aspekte menschlicher Intelligenz, inklusive Ich-Bewusstsein, innerem Erleben etc. steht dagegen nicht mehr im Mittelpunkt der Forschung.

Der Einsatz Künstlicher Intelligenz hat dennoch begonnen, verschiedenste Bereiche der Gesellschaft fundamental zu verändern. Insbesondere im Kontext der Wirtschaft wird dies überwiegend begrüßt bzw. hält man ihn für unvermeidlich, um im internationalen Wettbewerb auch künftig bestehen zu können. Die Bezeichnung «Industrie 4.0» steht für die Erwartung einer großen technischen Umwälzung, die vor allem durch die massive Vernetzung und Kollaboration von Maschinen und Menschen unter breitem Einsatz von KI gekennzeichnet ist.[3] Dies wirft grundsätzliche Fragen zur Arbeitswelt der Zukunft auf, die komplementär unter dem Stichwort «Arbeiten 4.0» diskutiert werden.[4] Bei vielen

1 John McCarthy, Marvin L. Minsky, Nathaniel Rochester, Claude E. Shannon: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, in: *AI Magazine*, Vol. 27, Nr 4, 2006 [1995], 12 f.

2 Ein Teil des Problems lag und liegt übrigens in den unscharfen Definitionen von «Intelligenz».

3 Vgl. Bundesministerium für Wirtschaft und Energie (Hg.): Was ist Industrie 4.0?, in: Plattform Industrie 4.0, *www.plattform-i40.de/I40/Navigation/DE/Industrie40/WasIndustrie40/was-ist-industrie-40.html*, gesehen am 7.1.2019.

4 Vgl. Bundesministerium für Arbeit und Soziales (Hg.): *Arbeit Weiter Denken. Weißbuch Arbeiten* 4.0, Berlin 2017, online unter *www.bmas.de/SharedDocs/Downloads/DE/PDF-Publikationen/a883-weissbuch.pdf?__blob=publicationFile*, gesehen am 8.4.2019.

Anwendungen geht es darum, KI in Form von künstlichen neuronalen Netzen (KNN) zu nutzen, um die infolge der Digitalisierung nahezu aller Lebensbereiche ungeheuer angewachsene Menge an verfügbaren Daten über menschliches Verhalten zu analysieren, Muster zu erkennen und diese zur Grundlage von Entscheidungen zu machen.

Zwar sind KNN nur eines von vielen Verfahren zur automatischen Entscheidungsfindung (Mustererkennung bzw. Klassifikation), aber aufgrund ihrer (erst) vor dem Hintergrund großer Datenmengen zustande kommenden Leistungsfähigkeit gelten sie im Moment als das am meisten Versprechende. KNN basieren auf einer komplexen Zusammenschaltung von sogenannten künstlichen Neuronen. Wie beim natürlichen Vorbild werden Informationen von Neuron zu Neuron über eine riesige Anzahl von Verbindungen weitergeleitet. Dabei werden sie durch sogenannte Netzgewichte jeweils graduell bestärkt oder negiert. Die für korrekte Entscheidungen optimalen Netzgewichte werden mithilfe von Maschinenlernverfahren (speziell dem sogenannten Fehlerrückverfolgungsverfahren) automatisch eingestellt. Eine typische Topologie für KNN ist eine schichtweise Anordnung, wobei auf den ‹unteren› Schichten Informationen von geringem Abstraktionsgrad und auf den ‹höheren› Schichten Informationen von großem Abstraktionsgrad bis hin zur finalen Entscheidung verarbeitet werden. Diese Schichtentopologie wird als mehrschichtiges Perzeptron (*multilayer perceptron*) bezeichnet. Existieren viele Schichten (etwa fünf oder mehr), spricht man von einem tiefen neuronalen Netz (*deep neural network*). Das entsprechende Maschinenlernverfahren heißt tiefes Lernen (*deep learning*).[5]

Die Einsatzmöglichkeiten von KNN, gerade auch bei der Analyse menschlichen Verhaltens, erscheinen unbegrenzt. Es kann um die Gewährung eines Kredites gehen,[6] um den möglichst effizienten Einsatz von Polizeistreifen zur Verhinderung von Verbrechen[7] oder um das Löschen von Hassbotschaften in sozialen Netzwerken.[8] Aufgrund der erheblichen Auswirkungen, die solche Anwendungen auf das Leben von Menschen haben können, sieht die 2018 in Kraft getretene Datenschutz-Grundverordnung der EU (DSGVO) vor, dass Personen

> das Recht haben, keiner Entscheidung […] zur Bewertung von sie betreffenden persönlichen Aspekten unterworfen zu werden, die ausschließlich auf einer automatisierten Verarbeitung beruht und die rechtliche Wirkung für die betroffene Person entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.[9]

Zwar besteht dieses Recht nicht, wenn eine bestimmte automatisierte Verarbeitung nach nationalem oder EU-Recht ausdrücklich erlaubt ist. Allerdings sieht in diesen Fällen die DSGVO u. a. den Anspruch der betroffenen Personen «auf Erläuterung der nach einer entsprechenden Bewertung getroffenen Entscheidung»[10] vor. Ein solcher Anspruch ist jedoch alles andere als trivial: Werden maschinelle Entscheidungen auf Basis eines KNN getroffen, so sind sie für Menschen nicht ‹nachvollziehbar›. Die Prozesse neuronaler

5 Es existieren weitere wichtige Netzwerktopologien, z. B. Netze mit Rückkopplungen oder bidirektionalem Informationsfluss, sowie verfeinerte Neuronenmodelle, z. B. sogenannte *long short-term memories*. Mithilfe speziell vereinfachter Neuronen können elementare Filter für Töne und Bilder realisiert werden, die der Informationsvorverarbeitung des Ohres und des Auges ähnlich sind (konvolutive neuronale Netze).

6 Vgl. Amir E. Khandani, Adlar J. Kim, Andrew W. Lo: Consumer Credit-Risk Models via Machine-Learning Algorithms, in: *Journal of Banking & Finance*, Vol. 34, Nr. 11, 2767–2787.

7 Vgl. Claudia Aradau, Tobias Blanke: Politics of Prediction. Security and the Time/Space of Governmentality in the Age of Big Data, in: *European Journal of Social Theory*, Vol. 20, Nr. 3, 373–391.

8 Vgl. Thomas Davidson, Dana Warmsley, Michael Macy u. a.: Automated Hate Speech Detection and the Problem of Offensive Language, in: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, Palo Alto 2017, 512–515.

9 Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), in: *Amtsblatt der Europäischen Union* L 119 vom 4.5.2016, 1–88, hier Erwägungsgrund 71, online unter data.europa.eu/eli/reg/2016/679/oj, gesehen am 19.4.2019.

10 Ebd.

Netze sind zwar gänzlich nachrechenbar, aber weder formal-semantisch noch intuitiv interpretierbar, da die immense Vielzahl von Neuronenverbindungen für Menschen absolut unüberblickbar ist. Es ist demnach zu Recht von ‹Black-Box-Verfahren›[11] die Rede.

II. Entscheidungsrationalität und Uninterpretierbarkeit

Die Intransparenz der von KI-Systemen getroffenen Entscheidungen kontrastiert auf den ersten Blick stark mit Entscheidungen, die von Menschen gefällt werden. Menschen haben normalerweise keine Schwierigkeiten damit, eine Erklärung für ihre Handlungen oder Entscheidungen zu geben. Ob und wie viel eine solche Erklärung damit zu tun hat, wie eine Handlung oder Entscheidung tatsächlich zustande gekommen ist, ist allerdings eine ganz andere Frage. Motive, auf die Menschen in alltagsweltlichen Kontexten verweisen, um ihr Verhalten zu erläutern, sind nicht Ursache oder Anstoß des Handelns, sondern Vokabulare, die in bestimmten Situationen akzeptiert sind, wenn es darum geht, Verbindungen zwischen Handlungen und wahrscheinlichen Konsequenzen dieser Handlungen herzustellen.[12] Die Rationalität menschlichen Entscheidens ist dadurch eingeschränkt, dass Menschen in Entscheidungssituationen nur begrenzte Informationen zur Verfügung haben und ihre Kapazitäten, diese zu verarbeiten sowie die Konsequenzen unterschiedlicher Optionen zu erwägen, ebenfalls begrenzt sind.[13] Sie entscheiden oft auf der Basis von Intuition und Emotionen ohne sorgfältige Analyse und sind insofern anfällig für unterschiedliche Arten kognitiver Verzerrungen.[14]

KI-Anwendungen versprechen hingegen, die kognitiven Grenzen der Informationsverarbeitung so weit zu verschieben, dass sie alle irgendwie verfügbaren Informationen tatsächlich einbeziehen können. Die damit in Aussicht gestellte gesteigerte Entscheidungsrationalität[15] wird aber gleichzeitig dadurch in Frage gestellt, dass der tatsächliche algorithmenbasierte Prozess genau wie jede menschliche Entscheidung durch Abstraktion, also Informationsverlust, charakterisiert ist. Ein automatischer Entscheidungsprozess ist zwar frei von kognitiven Verzerrungen, jedoch im Allgemeinen nicht frei von ‹Vorurteilen›. Beispielsweise erwartet ein Schriftzeichenerkenner wesentlich stärker, den Buchstaben E zu erkennen, als den Buchstaben X, was rational durch die viel größere Häufigkeit des Ersteren, die sogenannte A-priori-Wahrscheinlichkeit, in der deutschen Schriftsprache begründet ist. Genau dieses begründete Vorurteil kann aber im Einzelfall zu Fehlentscheidungen führen.

Da solche objektiven Fehlentscheidungen begründet sind, treffen KI-Anwendungen, die auf Prinzipien des maschinellen Lernens beruhen,[16] ausschließlich rationale Entscheidungen im Sinne praktischer – am Verhältnis von Zweck und Mittel orientierter – wie formaler – d. h. auf universalen Regeln basierender – Rationalität.[17] Gleichzeitig kann die Abstraktionsleistung, die algorithmenbasierte Prozesse vollziehen, in der Gesellschaft verankerte Vorurteile

11 Im Gegensatz zu den sogenannten Black-Box-Verfahren (worunter KNN fallen, aber auch Hidden-Markov-Modelle) stehen White-Box-Verfahren. Hierbei handelt es sich beispielsweise um symbolische KIs, die auf Logikkalkülen basieren.

12 Vgl. C. Wright Mills: Situated Actions and Vocabularies of Motive, in: *American Sociological Review*, Vol. 5, Nr. 6, 1940, 904–913.

13 Vgl. Herbert A. Simon: *Models of Man, Social and Rational. Mathematical Essays on Rational Human Behavior in a Social Setting*, New York 1957.

14 Vgl. Daniel Kahneman: Maps of Bounded Rationality. Psychology for Behavioral Economics, in: *American Economic Review*, Vol. 93, Nr. 5, 2003, 1449–1475.

15 Vgl. Amanda Clarke, Jonathan Craft: The Vestiges and Vanguards of Policy Design in a Digital Context, in: *Canadian Public Administration*, Vol. 60, Nr. 4, 2017, 476–497.

16 Man unterscheidet dabei im Wesentlichen überwachtes, unüberwachtes und Verstärkungslernen. Bei Ersterem wird dem Lernalgorithmus zu allen Lernbeispielen ein Gegenstück zur Verfügung gestellt, z. B. Fotos von Gesichtern und Namen der betreffenden Personen. Bei den anderen beiden Paradigmen ist dies nicht der Fall. Dort muss der Lernalgorithmus also z. B. ähnliche Fotos automatisch als zu ein und derselben Person zugehörig einstufen. Beim Verstärkungslernen bekommt das System nach einer Entscheidung (z. B. «Das Bild zeigt Person XY») eine Rückmeldung, im einfachsten Fall «richtig» oder «falsch», und kann mit dieser Information zukünftige Entscheidungen verbessern. Viele KI-Systeme benutzen eine Kombination mehrerer Lernmethoden. Die Aufgabe von maschinellem Lernen besteht demnach darin, Datenmodelle, die Grundlage maschineller Entscheidungen und Handlungen, automatisch aus Lernbeispielen aufzubauen.

17 Zu diesen und weiteren Typen von Rationalität im Anschluss an Max Weber vgl. Stephen Kalberg: Max Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History, in: *American Journal of Sociology*, Vol. 85, Nr. 5, 1980, 1145–1179.

nicht nur reproduzieren, sondern sogar verstärken. Auf menschliches Verhalten angewandt, entfaltet die Erkennung von Mustern einen performativen Effekt, der aufgefundene Ähnlichkeiten und Differenzen naturalisiert.[18]

Insofern moderne Gesellschaften von Rationalitätserwartungen durchdrungen sind, die zweckrationale Kalkulation auf der Basis allgemeiner Regeln als Ausweis von Handlungsfähigkeit prämiieren,[19] stellen KI-Anwendungen trotz dieser Probleme ein potenzielles Ideal für Entscheidungen dar. Sie bieten sich als Mittel für verschiedenste Zwecke an und versprechen, strikt regelgeleitet auf der Basis von Algorithmen rationale Entscheidungen zu treffen.

Bezogen auf KNN entfaltet dieses Versprechen allerdings nur dann seine volle Suggestionskraft, wenn der Fokus auf der Berechenbarkeit und Reproduzierbarkeit der Entscheidungen liegt und nicht auf ihrer Uninterpretierbarkeit. Verfahren der *explainable artificial intelligence* haben das Ziel, Entscheidungen von KI-Systemen für Menschen nachvollziehbar zu machen, bislang allerdings mit begrenztem Erfolg.[20] Eines der hauptsächlichen technischen Probleme liegt darin, dass *explainable AI* im Wesentlichen nur Einzelentscheidungen erklären kann und aufgrund einer zu wenig ausgereiften Semantikmodellierung nicht in der Lage ist, komplexe Situationen zu erfassen und zu interpretieren.

Eine wegweisende technische Entwicklung in diese Richtung findet derzeit auf dem Gebiet der kognitiven Prüftechnik bzw. Materialdiagnostik statt. Zwar befasst sich die aktuelle Forschung lediglich mit der KI-basierten Prüfung von Bauteilen, Maschinen und Anlagen, doch kann das Prüfprinzip auch auf die KI selbst ausgeweitet werden. Der Nutzen ist offensichtlich: Menschen können Entscheidungen von Black-Box-Verfahren nicht durchschauen, KI-Systeme hingegen können dies durchaus, was allerdings nicht automatisch mit Transparenz einhergeht, weder hinsichtlich des geprüften noch des prüfenden KI-Systems. Das kognitive Prüfsystem muss seine Entscheidungen und Ergebnisse dem Menschen nachvollziehbar machen können. Technologisch kommen also entweder White-Box-Verfahren[21] oder kognitive Mensch-Maschine-Schnittstellen in Frage.[22] Erstere liefern konstruktionsbedingt nachvollziehbare Entscheidungen, Letztere können beispielsweise durch Ermöglichung *natürlichsprachiger* Kommunikation mit Black-Box-Systemen[23] komplexe Sachverhalte erklären und auf gezielte Rückfragen seitens des Menschen aussagekräftige Antworten geben.[24]

Zwar sind aktuelle Sprachassistenten noch nicht leistungsfähig genug für eine komplexe Kommunikation zwischen Mensch und Maschine, der Hauptgrund dafür ist jedoch die nur rudimentär vorhandene rechentechnische Verarbeitung von Sprach- und Situationsbedeutungen. Für die Entwicklung einer funktionalen und effizienten rechentechnischen Prozessierung von Kommunikation ist die Expertise der Medienwissenschaften nicht nur gefragt, sondern auch notwendig. Die in der Medienwissenschaft und in ihren ‹Vorgängerwissenschaften› (wie z. B. der Literaturwissenschaft) entwickelten (im Vergleich zur Linguistik auf einer Meso- oder Makro-Ebene ansetzenden) Verfahren der Text- und

18 Vgl. Wendy Hui Kyong Chun: Queerying Homophily. Muster der Netzwerkanalyse, in: *Zeitschrift für Medienwissenschaft*, Nr. 18, H. 1, 2018, 131–148.

19 Vgl. Kalberg: Max Weber's Types of Rationality, 1158; John W. Meyer, Ronald L. Jepperson: The «Actors» of Modern Society. The Cultural Construction of Social Agency, in: *Sociological Theory*, Vol. 18, Nr. 1, 2000, 100–120; Thomas Lemke: Neoliberalismus, Staat und Selbsttechnologien. Ein kritischer Überblick über die ‹governmentaliy studies›, in: *Politische Vierteljahresschrift*, Bd. 41, Nr. 1, 2000, 31–47.

20 Vgl. für einen schnellen (und gut verständlichen) Überblick Paul Voosen: How AI Detectives Are Cracking Open the Black Box of Deep Learning, in: *ScienceMag.org*, dort datiert 6.7.2017, DOI: 10.1126/science.aan7059, gesehen am 7.1.2019, und aus medienwissenschaftlicher Perspektive Andreas Sudmann: On the Media-political Dimension of Artificial Intelligence. Deep Learning as a Black Box and OpenAI, in: *Digital Culture & Society*, Vol. 4, Nr. 1, 2018, 181–200.

21 Vgl. Anm. 11.

22 Die Frage, ob unabhängige KI-basierte Prüfsysteme für KI-Systeme eingeführt werden oder ob die Prüfung und Selbstinterpretation von vornherein in sicherheits- und rechtsrelevante KI-Systeme integriert wird, spielt für das Prinzip der Prüfung und Überwachung von KI durch KI nur eine untergeordnete Rolle.

23 Vgl. Upol Ehsan, Brent Harrison, Larry Chan u. a.: Rationalization: A Neural Machine Translation Approach to Generating Natural Language Explanations, in: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, New York 2018, 81–87.

24 Die stetige Verbesserung der Mensch-Maschine-Schnittstelle war (und ist) auch jenseits der KI-Systeme für die gesellschaftliche Akzeptanz von Computertechnik von größter Bedeutung. Man denke nur an die grafischen Benutzeroberflächen bei Personal Computern.

Kulturinterpretation lassen sich, eine Formalisierung bzw. Mathematisierung vorausgesetzt, gewinnbringend in die technische Entwicklung einbringen.[25] Die Medienwissenschaft könnte so genau jene Lücke schließen, die die Linguistik in Form der Computerlinguistik nicht zu schließen vermag.[26] Unabhängig davon, wie die technische Realisierung zustande kommt, könnte eine Überwachung von KI durch KI, bzw. exakter: der Einsatz von KI als Medium zwischen KI und Mensch, den Rückgewinn der Kontrolle des Menschen über die (kognitive) Maschine bedeuten. Das könnte manches ethische und gesellschaftspolitische Dilemma lösen – vorausgesetzt, dass die Auseinandersetzung mit KI im Allgemeinen und der Kontrolle von KI mittels KI im Besonderen von den Sozial-, aber eben auch und insbesondere den Medienwissenschaften kritisch begleitet, reflektiert und gegebenenfalls hinterfragt wird.

III. Kontrollgewinne durch Organisationsbildung

Die Wahrnehmung eines Kontrollverlustes im Zuge der Einführung neuer Technologien ist prinzipiell ein vertrautes Phänomen. Die Dampfmaschine als paradigmatische technologische Neuerung der ersten industriellen Revolution bietet hier ein instruktives Beispiel. Nach einer initialen Phase ihrer Erfindung und Entwicklung folgte eine Phase der rasanten Weiterentwicklung zu immer leistungsfähigeren Maschinen sowie deren massenhafter Einsatz. Dies führte aufgrund der noch nicht ausgereiften Technik zu teils lebensgefährlichen Dampfkesselexplosionen und resultierte so in einer (soziotechnischen) Kontrollverlustphase.

In Deutschland waren es daraufhin zunächst Ministerialbeamte der deutschen Staaten, die über Installationslizenzen entschieden und auf der Basis von ad hoc eingeholter Expertise Sicherheitsstandards festlegten – mit dem Ergebnis extrem rigider Konstruktionsvorgaben, die technische Weiterentwicklungen verhinderten.[27] Infolgedessen schlossen sich die Kesselbetreiber_innen zu freiwilligen Dampfkessel-Überwachungs- und Revisionsvereinen zusammen, die sehr erfolgreich bei der Verhütung von Dampfkesselunfällen waren. Aus diesen Vereinen gingen später die Technischen Überwachungsvereine (TÜV) hervor, die eine generelle Kontrollgewinnphase einleiteten und bis heute erheblich zur technischen Sicherheit von Anlagen, Maschinen und Fahrzeugen beitragen.[28]

Dieser kurze Blick in die Technikgeschichte liefert zwei wichtige Anregungen für das Nachdenken über KI-Anwendungen. Erstens zeigt er, dass nicht staatliche Organisationen eine wichtige Rolle in dem Bemühen spielen, Kontrollverluste in Kontrollgewinne zu überführen, und gesetzlichen Regelungen vorgreifen, sie ergänzen oder konkretisieren können. Staatliche Organisationen können natürlich ebenfalls einen wichtigen Beitrag leisten, insbesondere durch die Beratung politischer Entscheider_innen.[29] Aber Technikentwicklung und -anwendung finden maßgeblich im Kontext von privaten Organisationen statt,

25 Vgl. die modallogische Formalisierung von Jurij M. Lotmans Grenzüberschreitungstheorie (Peter Klimczak: *Formale Subtextanalyse. Kalkülisierung von Narration und Interpretation*, Münster 2016) und deren informationstechnische Umsetzbarkeit (Peter Klimczak, Petra Hofstedt, Ingo Schmitt u. a.: Computergestützte Methoden der Interpretation: Perspektiven einer digitalen Medienwissenschaft, in: Manuel Burghardt, Claudia Müller-Birn [Hg.]: *INF-DH 2018*, Bonn 2018, online unter dx.doi.org/10.18420/infdh2018-06, gesehen am 22.04.2019).

26 Vgl. Henning Lobin: Sprachautomaten, Eintrag im Blog *Die Engelbart-Galaxis. Digitale Welten jenseits der Schriftkultur*, dort datiert 25.5.2017, scilogs.spektrum.de/engelbart-galaxis/sprachautomaten/, gesehen am 17.4.2019.

27 Vgl. Peter Lundgreen: Scientific Expertise and Regulatory Politics in Germany. The Formative Period of Handling Risks by Agreeing on «Acceptable» Standards, 1870–1913, in: *1996 International Symposium on Technology and Society. Technical Expertise and Public Decision. Proceedings*, Princeton, New Jersey 1996, 532–536.

28 Vgl. Frank Uekötter: Entstehung des TÜV, in: Armin Grunwald, Melanie Simonidis-Puschmann (Hg.): *Handbuch Technikethik*, Stuttgart 2013, 50–55.

29 Vgl. z. B. das Büro für Technikfolgenabschätzung beim Deutschen Bundestag und dessen laufende Untersuchung: Mögliche Diskriminierung durch algorithmische Entscheidungsprozesse und maschinelles Lernen, last update 2.4.2019, www.tab-beim-bundestag.de/de/untersuchungen/u40400.html, gesehen am 8.4.2019.

woraus sich sowohl deren Interesse an Einfluss auf Regulierungsversuche ergibt als auch ein Informationsvorsprung, der die Berücksichtigung dieses Interesses nahezu unvermeidlich macht.[30]

Zweitens lässt sich zwar retrospektiv von einer ausgereiften Technik sprechen, aber die Etablierung einer Technik ist ein kontingenter sozialer Prozess, an dem neben den eigentlichen Entwickler_innen verschiedene Akteur_innen beteiligt sind.[31] Dabei ist die Frage der Kontrolle über technische Systeme eingebettet in einen Kontext, in dem Organisationsbildung und Professionalisierung nicht nur dazu beitragen, technische Standards zu entwickeln und durchzusetzen, sondern auch dazu, die soziale Akzeptanz einer neuen Technik zu erhöhen.[32] Dies wiederum ist entscheidend für die Akzeptanz der Organisationen, die diese Technik verwenden.

IV. Soziale Akzeptanz als Herausforderung

Organisationen benötigen für ihren Fortbestand nicht nur materielle Ressourcen und Informationen über Kund_innen oder Klient_innen und andere Aspekte ihrer Umwelt; sie brauchen Legitimität. Damit ist eine allgemeine Wahrnehmung oder Einschätzung gemeint, nach der die Aktivitäten einer Organisation, gemessen an akzeptierten Normen, Werten und Überzeugungen, wünschenswert oder angemessen sind.[33] Insofern KI in organisatorische Abläufe und Entscheidungsprozesse eingebunden wird, sind technische und organisatorische Lösungen, die Kontrollgewinne hinsichtlich der neuen Technologie ermöglichen, untrennbar mit der Legitimität der sie verwendenden und kontrollierenden Organisationen verknüpft.

Organisationen hängen insbesondere davon ab, dass sie ihr Handeln gegenüber ihrer sozialen Umwelt als rational darstellen können.[34] Die Legitimation des Einsatzes von KI kann davon dank des verbreiteten Mythos vom Computer als Verkörperung des Rationalitätsideals[35] vermutlich profitieren, insbesondere solange das Wissen über die Funktionsweise maschinellen Lernens in Politik und Öffentlichkeit relativ gering ist. Darüber hinaus spielen staatliche Initiativen wie «Plattform Industrie 4.0»[36] eine zentrale Rolle, weil sie einen organisationsförmigen Kontext von Arbeitsgruppen schaffen, in dem die Erwartung, dass KI-Anwendungen künftig nicht nur nützlich, sondern geradezu unabdingbar sein werden, als unhinterfragte Prämisse institutionalisiert ist.

Die Beteiligung von Expert_innen aus Unternehmen, Verbänden, Betriebsräten und Wissenschaft an den Arbeitsgruppen folgt dabei einem Muster, das in jüngerer Zeit insbesondere mit dem Stichwort Governance verbunden, im Rahmen von neokorporatistischen Arrangements aber schon wesentlich länger etabliert ist: Die Verflechtung staatlicher und nicht staatlicher Organisationen beim Umgang mit komplexen gesellschaftlichen Herausforderungen gilt nicht nur als unvermeidlich, sondern sogar als wünschenswert.[37] Ein solches Vorgehen hat stets auch Besorgnis und Kritik hervorgerufen, weil es privaten

30 Vgl. Arie Rip: *Futures of Science and Technology in Society*, Wiesbaden 2018, Kap. 4.

31 Vgl. Trevor J. Pinch, Wiebe E. Bijker: The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other, in: *Social Studies of Science*, Vol. 14, Nr. 3, 1984, 399–441.

32 Vgl. Mark C. Suchman: Managing Legitimacy. Strategic and Institutional Approaches, in: *The Academy of Management Review*, Vol. 20, Nr. 3, 1995, 571–610.

33 Vgl. ebd., 574.

34 Vgl. John W. Meyer, Brian Rowan: Institutionalized Organizations: Formal Structure as Myth and Ceremony, in: *American Journal of Sociology*, Vol. 83, Nr. 2, 1977, 340–363.

35 Vgl. Stefan Kuhlmann: Computer als Mythos, in: *Technik und Gesellschaft. Jahrbuch*, Bd. 3, 1985, 91–106; Reinhard Bahnmüller, Michael Faust: *Das automatisierte Arbeitsamt. Legitimationsprobleme, EDV-Mythos und Wirkungen des Technikeinsatzes*, Frankfurt / M., New York 1992.

36 Vgl. Bundesministerium für Wirtschaft und Energie (Hg.): Industrie 4.0.

37 Vgl. Jan Kooiman: *Governing as Governance*, Thousand Oaks 2003; Philippe C. Schmitter: Neo-Corporatism, in: Bertrand Badie, Dirk Berg-Schlosser, Leonardo Morlino (Hg.): *International Encyclopedia of Political Science*, Vol. 5, Thousand Oaks 2011, 1669–1673.

Interessen erheblichen Einfluss auf grundlegende Fragen politischer Gestaltung einräumt. Diese Problematik besteht natürlich auch im Zusammenhang mit KI, erst recht angesichts der Tatsache, dass sich die ursprüngliche Euphorie über die Möglichkeiten digitaler Vernetzung gerade bei einigen Pionier_innen des Web 2.0 inzwischen in Skepsis und Ablehnung verwandelt hat.[38] Diese speisen sich allerdings bislang überwiegend daraus, dass Anwendungen maschinellen Lernens das Potenzial zugeschrieben wird, tatsächlich Entscheidungen zu treffen, die das Verhalten jedes_r Einzelnen maßgeschneidert durch die richtige Botschaft zum richtigen Zeitpunkt beeinflussen – im Interesse des Profits von Unternehmen oder im Interesse des Erreichens anderer Ziele, die Organisationen verfolgen mögen.[39] (Berechtigter) Fokus der Kritik ist hier nicht die Technologie als solche, sondern es sind die Interessen, denen sie dient, und die Organisationen – seien es private Firmen oder auch Staaten –, die sich einseitig an diesen Interessen orientieren. Gleichzeitig konzentriert sich diese in der Öffentlichkeit wahrnehmbare Kritik gerade auf die vermeintlich überlegene und insofern bedrohliche Rationalität von KI-Anwendungen und stützt auf diese Weise sogar die Erwartung überlegener Entscheidungsrationalität, die sich wiederum in die allgemeinen Rationalitätserwartungen moderner Gesellschaften fügt.[40]

Die Entwicklung von Prüf- und Zertifizierungsvorgängen für die Überwachung von KI adressiert vor diesem Hintergrund ein einerseits kleines und andererseits großes Problem. Klein erscheint es in Relation zum Gesamtkomplex der Akzeptanzfrage, die eben keineswegs nur von der Kontrolle über die Technologie selbst abhängt. Gleichzeitig ist es aber groß, weil, unabhängig davon, welche Organisationen KI für wie immer gemeinwohlorientierte Ziele einsetzen wollen, soziale Akzeptanz für KI langfristig kaum ohne Kontrollgewinne hinsichtlich der Nachvollziehbarkeit durch sie getroffener Entscheidungen zu haben sein dürfte. Nicht staatliche Professionalisierungs- und Zertifizierungsprozesse, die einen Stand der Technik hinsichtlich der Kontrolle und Prüfung von KI etablieren und sich dabei an vertrauten Ideen wie Auditierung und Standardisierung orientieren,[41] fördern die Legitimität der Technologie und der sie verwendenden Organisationen. Auf diese Weise könnten KI-Anwendungen in naher Zukunft schließlich selbst *eine* Quelle der Legitimität von Organisationen und zum selbstverständlichen, als unverzichtbar betrachteten Element organisationalen Handelns werden.

38 Vgl. Jaron Lanier: *Ten Arguments for Deleting Your Social Media Accounts Right Now*, London 2018.

39 Vgl. Shoshana Zuboff: *The Age of Surveillance Capitalism. The Fight for the Future at the New Frontier of Power*, London 2019; Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer u. a.: Will Democracy Survive Big Data and Artificial Intelligence?, in: *Scientific American*, dort datiert 25.2.2017, *www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence*, gesehen am 7.1.2019.

40 Vgl. Meyer u. a.: The «Actors»; ders. u. a.: Institutionalized Organizations.

41 Vgl. Joshua A. Kroll: The Fallacy of Inscrutability, in: *Philosophical Transactions of the Royal Society A*, Vol. 376, Nr. 2133, 2018, o. S.